

CLAIMS:

1. A method for managing data comprising:
producing a probabilistically unique identifier for
a digital sequence; and
5 comparing said probabilistically unique identifier
to a list of other identifiers with their corresponding
digital sequences.
2. The method of claim 1 further comprising:
adding said probabilistically unique identifier to
10 said list if said probabilistically unique identifier is
not previously in said list.
3. The method of claim 1 further comprising:
removing said probabilistically unique identifier
from said list if said probabilistically unique
15 identifier is previously in said list.
4. The method of claim 2 further comprising:
adding said digital sequence corresponding to said
probabilistically unique identifier to said list.
5. The method of claim 3 further comprising:
20 removing said digital sequence corresponding to said
probabilistically unique identifier from said list.
6. The method of claim 4 further comprising:
adding a correspondence between said digital
sequence and said probabilistically unique identifier for
25 that sequence.
7. The method of claim 1 wherein said step of producing

comprises:

hashing said digital sequence to produce said
probabalistically unique identifier.

8. The method of claim 7 wherein said step of hashing
5 is carried out by means of an industry standard digest
algorithm.

9. The method of claim 8 wherein said step of hashing
is carried out by one of an MD4, MD5, SHA or SHA-1
algorithm.

10 10. The method of claim 1 wherein said step of producing
comprises:

generating a checksum for said digital sequence to
produce said probabilistically unique identifier.

11. The method of claim 1 wherein said digital sequence
15 is descriptive meta data of at least one other digital
sequence.

12. The method of claim 1 wherein said digital sequence
is descriptive meta data of at least one probabilistically
unique identifier.

20 13. The method of claim 1 wherein said digital sequence
describes a method that represents at least one digital
sequence.

14. A method for managing data comprising:
dividing a digital sequence into a plurality of
25 shorter digital sequences; and
producing probabilistically unique identifiers for

each said plurality of shorter digital sequences; and
comparing said probabilistically unique identifiers
to a list of other identifiers.

15. The method of claim 14 further comprising the step
5 of:

dividing said digital sequence into a plurality of
shorter digital sequences; and

producing a like plurality of probabilistically
unique identifiers corresponding to each of said
10 plurality of shorter digital sequences.

16. The method of claim 14 further comprising;
comparing each plurality of identifiers to said
list.

17. The method of claim 14 wherein said step of dividing
15 produces said shorter digital sequences having
individually variable lengths.

18. The method of claim 14 wherein said step of dividing
is based on the content of said digital sequence.

19. The method of claim 14 wherein said step of dividing
20 is based on meta data describing said digital sequence.

20. The method of claim 14 wherein said step of dividing
produces said shorter digital sequences having
substantially invariable lengths.

21. The method of claim 14 wherein said step of producing
25 said like plurality of probabilistically unique
identifiers comprises:

individually hashing said shorter digital sequences to produce said like plurality of probabilistically unique identifiers.

22. The method of claim 14 further comprising the step
5 of:

adding said plurality of shorter digital sequences and said corresponding like plurality of probabilistically unique identifiers to said list.

23. The method of claim 14 further comprising the step
10 of:

removing said plurality of shorter digital sequences and said corresponding like plurality of probabilistically unique identifiers from said list.

24. The method of claim 9 further comprising the step of:
15 utilizing at least a portion of said probabilistically unique identifier as an indicator to a location in said list for said step of comparing.

25. A computer program product comprising:
a computer usable medium having computer readable
20 code embodied therein for managing data, said computer program product comprising:

computer readable program code devices configured to cause a computer to effect producing a probabilistically unique identifier for a digital sequence; and
25 computer readable program code devices configured to cause a computer to effect comparing said probabilistically unique identifier to a list of other identifiers corresponding to other digital sequences.

26. The computer program product of claim 25 further comprising:

computer readable program code devices configured to cause a computer to effect adding said probabilistically
5 unique identifier to said list if said probabilistically unique identifier is not previously in said list.

27. The computer program product of claim 26 further comprising:

computer readable program code devices configured to
10 cause a computer to effect adding said corresponding digital sequence to said list.

28. The computer program product of claim 25 wherein said computer readable program code devices configured to cause said computer to effect producing comprises:

15 computer readable program code devices configured to cause a computer to effect hashing said digital sequence to produce said probabilistically unique identifier.

29. The computer program product of claim 28 wherein said computer readable program code devices configured to
20 cause a computer to effect hashing is carried out by means of an industry standard digest algorithm.

30. The computer program product of claim 29 wherein said computer readable program code devices configured to cause a computer to effect hashing is carried out by one
25 of an MD4, MD5, SHA or SHA-1 algorithm.

31. The computer program product of claim 25 wherein said computer readable program code devices configured to

cause a computer to effect producing comprises:

computer readable program code devices configured to
cause a computer to effect generating a checksum for said
digital sequence to produce said probabilistically unique
5 identifier.

32. The computer program product of claim 25 further
comprising:

computer readable program code devices configured to
cause a computer to effect creating a directory list
10 containing said probabilistically unique identifier for
said digital sequence.

33. The computer program product of claim 25 further
comprising:

computer readable program code devices configured to
15 cause a computer to effect dividing said digital sequence
into a plurality of shorter digital sequences; and

computer readable program code devices configured to
cause a computer to effect producing a like plurality of
probabilistically unique identifiers corresponding to
20 each of said plurality of shorter digital sequences.

34. The computer program product of claim 33 wherein said
computer readable program code devices configured to
cause a computer to effect dividing produces said shorter
digital sequences having individually variable length.

25 35. The computer program product of claim 33 wherein said
computer readable program code devices configured to
cause a computer to effect dividing produces said shorter
digital sequences having substantially invariable

length.

36. The computer program product of claim 33 wherein said computer readable program code devices configured to cause a computer to effect producing said like plurality
5 of probabilistically unique identifiers comprises:

computer readable program code devices configured to cause a computer to effect individually hashing said shorter digital sequences to produce said like plurality of probabilistically unique identifiers.

10 37. The computer program product of claim 33 further comprising:

computer readable program code devices configured to cause a computer to effect adding said plurality of shorter digital sequences and said corresponding like
15 plurality of probabilistically unique identifiers to said list.

38. The computer program product of claim 25 further comprising:

computer readable program code devices configured to
20 cause a computer to effect utilizing at least a portion of said probabilistically unique identifier as an index into a table of locations for said list for said step of comparing.

39. A method for managing data comprising:

25 producing a probabilistically unique identifier for a digital sequence; and

comparing said probabilistically unique identifier to a list of other identifiers corresponding to other

digital sequences.

40. The method of claim 39 further comprising:

adding said probabilistically unique identifier to
said list if said probabilistically unique identifier is
5 not previously in said list.